GETTING STARTED WITH DIRECT MT

Milind Ganjoo
Outline

• Direct MT approach
• Adding transfer rules
• Analyzing word alignments
• Examples and inferences
Step 1: Dictionary translation

- One “foreign” word → one (possibly more) English dictionary word(s)
  - e.g. rouge → red, aux → to the
- Good starting point: [http://www.wordreference.com/](http://www.wordreference.com/)
- Problems with this approach?
  - Context
    - e.g. été means “summer” when used by itself, but is an auxiliary verb in the past tense of “to be” when used as “a été”
    - Ignore context for this exercise – pick the most common independent meaning
  - Tokenization
    - e.g. Qu’est-ce que tu fais? (French)
    - It’s ok to normalize morphemes:
    - Que, est, ce, que, tu, fais
Step 1: Dictionary translation

• One “foreign” word $\rightarrow$ one (possibly more) English dictionary word(s)
  • e.g. rouge $\rightarrow$ red, aux $\rightarrow$ to the

• Good starting point: http://www.wordreference.com/
Step 2: Reordering

- Use simple POS-based rules
- Rules should be *general* enough
  - e.g. noun adjective → adjective noun
- POS tags must be on *individual* words, bereft of context
Let’s work through an example!

- French: Elle aime la robe rouge.

- Find the direct translation

- English: She loves the red dress.
- Google Translate: She loves red dress.
Example 2

- French: Qu’est-ce que vous faites?
- English: What do you do? (crude translation)

- **Find the direct translation**

- It’s ok to normalize morphemes:
  - Que, est, ce, que, tu, fais
Example 3

- French: La lettre a été envoyée le Mardi
- Direct translation: The letter went was summer sent the Tuesday
- Actual: The letter was sent on Tuesday

**What does Google Translate say?**

- Context is important
  - e.g. été means “summer” when used by itself, but is an auxiliary verb in the past tense of “to be” when used as “a été”
  - Ignore context for this exercise – pick the most common independent meaning
Useful analysis tool: word alignment

a[i] 0 1 -1 3 2
f[i] il aime la musique classique

he likes classical music
Picaro: alignment visualization tool

- [http://www.isi.edu/~riesa/software/picaro/](http://www.isi.edu/~riesa/software/picaro/)
- Try the demo: “Web version”
- Use English as “e” language, and your chosen language as “f”
- Alignment input format: f-e
Picaro: alignment visualization tool

- Example
  - French: Il aime la musique classique
  - English: He loves classical music
- What would the alignment format be?
  - Il → he, aime → love, la → (), musique → music, classique → classical
- Alignment: 0-0 1-1 3-3 4-2
Outputting alignment from your code

Elle aime la robe rouge
She loves the dress red

Swap alignments

She loves the red dress
// translate
for each i:
    english_trans[i] = lookup_dictionary(french[i])
    alignment[i] = i

// reordering
// NN JJ -> JJ NN
for each i:
    if english_trans[i] is noun and english_trans[i+1] is adjective:
        swap(english_trans[i], english_trans[i+1])
        swap(alignment[i], alignment[i+1])
Let’s try another language – Hindi!

• One of the 22 official languages of India
• Regional dialect/vernacular developed in the first millenium
• Absorbed large number of Persian, Arabic and Turkish words (and English words later)
• Standardized by the Government of India in 1958
**A simple sentence**

- मैं दुकान से दूध खरीदने जा रहा हूँ।
- Direct translation:
  
  मैं दुकान से दूध खरीदने जा रहा हूँ
  
  I store from milk buy go is am

- Actual translation:
  - I am going to buy milk from the store
- Google Translate:
  - I'm going to buy milk from the store
- Language divergences:
  - SOV
  - “Postpositions” – “from” comes after noun “store”
  - Tenses for verbs require multiple auxiliary verbs
Other simple examples

• दुनिया में दो प्रकार के लोग होते हैं।
• Direct translation:
  • World in two type of people are are are
• Actual translation:
  • There are two types of people in the world
• Google Translate:
  • There are two types of people in the world
• Information about singular/plural nouns is not directly presented – determined from verb agreement
Other simple examples

• कत्त मंत्री ने आज नए बजट का ऐलान कया है।

• Direct translation:
  • Finance minister has today new budget of announcement made is

• Actual translation:
  • The Finance minister announced the new budget today

• Google Translate:
  • Finance Minister has announced today a new budget

• Hindi does not contain articles (a/an/the) – specificity is determined from context
Things don’t always work well…

• उसने अपने बच्चों को अपना घर और अपनी सारी दौलत दे दी

• Direct translation:
  • He gave his children his house and all his wealth

  • He gave his children his house and all his wealth
  • (Could be “she/her” as well)

• Google Translate:
  • He gave all your money to your children and your home.

  Can’t determine gender since we aren’t using
Try to obtain as much grammatical diversity as possible

See slides on Language Divergences for